

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Bandman, Olga
Corley, Neil C.
Guegler, Karl J.
Lal, Preeti
- (ii) TITLE OF THE INVENTION: PROSTATE-ASSOCIATED ETS PROTEIN
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
(B) STREET: 3174 Porter Dr.
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: To Be Assigned
(B) FILING DATE: Filed Herewith
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Billings, Lucy J.
(B) REGISTRATION NUMBER: 36,749
(C) REFERENCE/DOCKET NUMBER: PF-0501 US
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 650-855-0555
(B) TELEFAX: 650-845-4166
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 335 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
(A) LIBRARY: PROSTUT12
(B) CLONE: 1813005
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Gly Ser Ala Ser Pro Gly Leu Ser Ser Val Ser Pro Ser His Leu
1 5 10 15

Leu Leu Pro Pro Asp Thr Val Ser Arg Thr Gly Leu Glu Lys Ala Ala
 20 25 30
 Ala Gly Ala Val Gly Leu Glu Arg Arg Asp Trp Ser Pro Ser Pro Pro
 35 40 45
 Ala Thr Pro Glu Gln Gly Leu Ser Ala Phe Tyr Leu Ser Tyr Phe Asp
 50 55 60
 Met Leu Tyr Pro Glu Asp Ser Ser Trp Ala Ala Lys Ala Pro Gly Ala
 65 70 75 80
 Ser Ser Arg Glu Glu Pro Pro Glu Glu Pro Glu Gln Cys Pro Val Ile
 85 90 95
 Asp Ser Gln Ala Pro Ala Gly Ser Leu Asp Leu Val Pro Gly Gly Leu
 100 105 110
 Thr Leu Glu Glu His Ser Leu Glu Gln Val Gln Ser Met Val Val Gly
 115 120 125
 Glu Val Leu Lys Asp Ile Glu Thr Ala Cys Lys Leu Leu Asn Ile Thr
 130 135 140
 Ala Asp Pro Met Asp Trp Ser Pro Ser Asn Val Gln Lys Trp Leu Leu
 145 150 155 160
 Trp Thr Glu His Gln Tyr Arg Leu Pro Pro Met Gly Lys Ala Phe Gln
 165 170 175
 Glu Leu Ala Gly Lys Glu Leu Cys Ala Met Ser Glu Glu Gln Phe Arg
 180 185 190
 Gln Arg Ser Pro Leu Gly Gly Asp Val Leu His Ala His Leu Asp Ile
 195 200 205
 Trp Lys Ser Ala Ala Trp Met Lys Glu Arg Thr Ser Pro Gly Ala Ile
 210 215 220
 His Tyr Cys Ala Ser Thr Ser Glu Glu Ser Trp Thr Asp Ser Glu Val
 225 230 235 240
 Asp Ser Ser Cys Ser Gly Gln Pro Ile His Leu Trp Gln Phe Leu Lys
 245 250 255
 Glu Leu Leu Leu Lys Pro His Ser Tyr Gly Arg Phe Ile Arg Trp Leu
 260 265 270
 Asn Lys Glu Lys Gly Ile Phe Lys Ile Glu Asp Ser Ala Gln Val Ala
 275 280 285
 Arg Leu Trp Gly Ile Arg Lys Asn Arg Pro Ala Met Asn Tyr Asp Lys
 290 295 300
 Leu Ser Arg Ser Ile Arg Gln Tyr Tyr Lys Lys Gly Ile Ile Arg Lys
 305 310 315 320
 Pro Asp Ile Ser Gln Arg Leu Val Tyr Gln Phe Val His Pro Ile
 325 330 335

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1905 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSTUT12
- (B) CLONE: 1813005

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CTCGAGCCGC	GGCTGTCTGA	CTTCCTCCCA	GCACATTCTC	GCACTCTGCC	GTGTCCACAC	60
TGCCCCACAG	ACCCAGTCCT	CCAAGCCTGC	TGCCAGCTCC	CTGCAAGCCC	CTCAGGTTGG	120
GCCTTGCCAC	GGTGCCAGCA	GGCAGCCCTG	GGCTGGGGGT	AGGGGACTCC	CTACAGGCAC	180
GCAGCCCTGA	GACCTCAGAG	GGCCACCCCT	TGAGGGTGGC	CAGGCCCCCA	GTGGCCAACC	240
TGAGTGCTGC	CTCTGCCACC	AGCCCTGCTG	GCCCCCTGGT	CCGCTGGCCC	CCCAGATGCC	300
TGGCTGAGAC	ACGCCAGTGG	CCTCAGCTGC	CCACACCTCT	TCCCGGCCCC	TGAAGTTGGC	360
ACTGCAGCAG	ACAGCTCCCT	GGGCACCAGG	CAGCTAACAG	ACACAGCCGC	CAGCCCCAAC	420
AGCAGCGGCA	TGGGCAGCGC	CAGCCCGGGT	CTGAGCAGCG	TATCCCCCAG	CCACCTCCTG	480
CTGCCCCCGG	ACACGGTGTC	GCGGACAGGC	TTGGAGAAGG	CGGCAGCGGG	GGCAGTGGGT	540
CTCGAGAGAC	GGGACTGGAG	TCCCAGTCCA	CCCGCCACGC	CCGAGCAGGG	CCTGTCCGCC	600
TTCTACCTCT	CCTACTTTGA	CATGCTGTAC	CCTGAGGACA	GCAGCTGGGC	AGCCAAGGCC	660

```

CCTGGGGCCA GCAGTCGGGA GGAGCCACCT GAGGAGCCTG AGCAGTGCCC GGTCAATTGAC 720
AGCCAAGCCC CAGCGGGCAG CCTGGACTTG GTGCCCCGCG GGCTGACCTT GGAGGAGCAC 780
TCGCTGGAGC AGGTGCAGTC CATGGTGGTG GGCGAAGTGC TCAAGGACAT CGAGACGGCC 840
TGCAAGCTGC TCAACATCAC CGCAGATCCC ATGGACTGGA GCCCCAGCAA TGTGCAGAAG 900
TGGCTCCTGT GGACAGAGCA CCAATACCGG CTGCCCCCA TGGGCAAGGC CTTCCAGGAG 960
CTGGCGGGCA AGGAGCTGTG CGCCATGTCT GAGGAGCAGT TCCGCCAGCG CTCGCCCCTG 1020
GGTGGGGATG TGCTGCACGC CCACCTGGAC ATCTGGAAGT CAGCGGCCTG GATGAAAGAG 1080
CGGACTTCAC CTGGGGCGAT TCACTACTGT GCCTCGACCA GTGAGGAGAG CTGGACCGAC 1140
AGCGAGGTGG ACTCATCATG CTCCGGGCAG CCCATCCACC TGTGGCAGTT CCTCAAGGAG 1200
TTGCTACTCA AGCCCCACAG CTATGGCCGC TTCATTAGGT GGCTCAACAA GGAGAAGGGC 1260
ATCTTCAAAA TTGAGGACTC AGCCCAGGTG GCGCGGCTGT GGGGCATCCG CAAGAACCGT 1320
CCCGCCATGA ACTACGACAA GCTGAGCCGC TCCATCCGCC AGTATTACAA GAAGGGCATC 1380
ATCCGGAAGC CAGACATCTC CCAGCGCCTC GTCTACCAGT TCGTGCACCC CATCTGAGTG 1440
CCTGGCCCCAG GGCCTGAAAC CCGCCCTCAG GGGCCTCTCT CCTGCCTGCC CTGCCTCAGC 1500
CAGGCCCTGA GATGGGGGAA AACGGGCAGT CTGCTCTGCT GCTCTGACCT TCCAGAGCCC 1560
AAGGTCAGGG AGGGGCAACC AACTGCCCCA GGGGGATATG GGTCTCTTGG GGCCTTCGGG 1620
ACCCTGGGGC AGGGGTGCTT CCTCCTCAGG CCCAGCTGCT CCCCTGGAGG ACAGAGGGAG 1680
ACAGGGCTGC TCCCCAACAC CTGCCTCTGA CCCCAGCATT TCCAGAGCAG AGCCTACAGA 1740
AGGGCAGTGA CTCGACAAAG GCCACAGGCA GTCCAGGCCT CTCTCTGCTC CATCCCCCTG 1800
CCTCCCATTG TGCACCACAC CTGGCATGGT GCAGGGAGAC ATCTGCACCC CTGAGTTGGG 1860
CAGCCAGGAG TGCCCCCGGG AATGGATAAT AAAGATACTA GGCGC 1905

```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 114 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 157196

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Thr Asn Ala Ser Asn Gly Gly Thr Ala Thr Val Lys Arg Pro Asn Gly
1      5      10      15
Gly Arg Thr Gly Gly Gly Ser His Ile His Leu Trp Gln Phe Leu
20     25     30
Lys Glu Leu Leu Ala Ser Pro Gln Val Asn Gly Thr Ala Ile Arg Trp
35     40     45
Ile Asp Arg Ser Lys Gly Ile Phe Lys Ile Glu Asp Ser Val Arg Val
50     55     60
Ala Lys Leu Trp Gly Arg Lys Asn Arg Pro Ala Met Asn Tyr Asp
65     70     75     80
Lys Leu Ser Arg Ser Ile Arg Gln Tyr Tyr Lys Lys Gly Ile Met Lys
85     90     95
Lys Thr Glu Arg Ser Gln Arg Leu Val Tyr Gln Phe Cys His Pro Tyr
100    105    110
Ser Gln

```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 452 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 511283

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Ser	Glu	Thr	Pro	Ala	Gln	Cys	Ser	Ile	Lys	Gln	Glu	Arg	Ile	Ser
1				5					10					15	
Tyr	Thr	Pro	Pro	Glu	Ser	Pro	Val	Pro	Ser	Tyr	Ala	Ser	Ser	Thr	Pro
			20					25					30		
Leu	His	Val	Pro	Val	Pro	Arg	Ala	Leu	Arg	Met	Glu	Glu	Asp	Ser	Ile
		35					40					45			
Arg	Leu	Pro	Ala	His	Leu	Arg	Leu	Gln	Pro	Ile	Tyr	Trp	Ser	Arg	Asp
	50					55					60				
Asp	Val	Ala	Gln	Trp	Leu	Lys	Trp	Ala	Glu	Asn	Glu	Phe	Ser	Leu	Arg
65					70					75				80	
Pro	Ile	Asp	Ser	Asn	Thr	Phe	Glu	Met	Asn	Gly	Lys	Ala	Leu	Leu	Leu
				85					90					95	
Leu	Thr	Lys	Glu	Asp	Phe	Arg	Tyr	Arg	Ser	Pro	His	Ser	Gly	Asp	Val
			100					105						110	
Leu	Tyr	Glu	Leu	Leu	Gln	His	Ile	Leu	Lys	Gln	Arg	Lys	Pro	Arg	Ile
		115					120					125			
Leu	Phe	Ser	Pro	Phe	Phe	His	Pro	Gly	Asn	Ser	Ile	His	Thr	Gln	Pro
130						135					140				
Glu	Val	Ile	Leu	His	Gln	Asn	His	Glu	Glu	Asp	Asn	Cys	Val	Gln	Arg
145					150					155					160
Thr	Pro	Arg	Pro	Ser	Val	Asp	Asn	Val	His	His	Asn	Pro	Pro	Thr	Ile
				165					170					175	
Glu	Leu	Leu	His	Arg	Ser	Arg	Ser	Pro	Ile	Thr	Thr	Asn	His	Arg	Pro
			180					185					190		
Ser	Pro	Asp	Pro	Glu	Gln	Arg	Pro	Leu	Arg	Ser	Pro	Leu	Asp	Asn	Met
		195					200					205			
Ile	Arg	Arg	Leu	Ser	Pro	Ala	Glu	Arg	Ala	Gln	Gly	Pro	Arg	Pro	His
	210					215					220				
Gln	Glu	Asn	Asn	His	Gln	Glu	Ser	Tyr	Pro	Leu	Ser	Val	Ser	Pro	Met
225					230					235					240
Glu	Asn	Asn	His	Cys	Pro	Ala	Ser	Ser	Glu	Ser	His	Pro	Lys	Pro	Ser
				245					250					255	
Ser	Pro	Arg	Gln	Glu	Ser	Thr	Arg	Val	Ile	Gln	Leu	Met	Pro	Ser	Pro
			260					265					270		
Ile	Met	His	Pro	Leu	Ile	Leu	Asn	Pro	Arg	His	Ser	Val	Asp	Phe	Lys
		275					280					285			
Gln	Ser	Arg	Leu	Ser	Glu	Asp	Gly	Leu	His	Arg	Glu	Gly	Lys	Pro	Ile
		290				295					300				
Asn	Leu	Ser	His	Arg	Glu	Asp	Leu	Ala	Tyr	Met	Asn	His	Ile	Met	Val
305					310					315					320
Ser	Val	Ser	Pro	Pro	Glu	Glu	His	Ala	Met	Pro	Ile	Gly	Arg	Ile	Ala
				325					330					335	
Asp	Cys	Arg	Leu	Trp	Asp	Tyr	Val	Tyr	Gln	Leu	Leu	Ser	Asp	Ser	
			340				345					350			
Arg	Tyr	Glu	Asn	Phe	Ile	Arg	Trp	Glu	Asp	Lys	Glu	Ser	Lys	Ile	Phe
		355					360					365			
Arg	Ile	Val	Asp	Pro	Asn	Gly	Leu	Ala	Arg	Leu	Trp	Gly	Asn	His	Lys
		370				375					380				
Asn	Arg	Thr	Asn	Met	Thr	Tyr	Glu	Lys	Met	Ser	Arg	Ala	Leu	Arg	His
385					390					395					400
Tyr	Tyr	Lys	Leu	Asn	Ile	Ile	Arg	Lys	Glu	Pro	Gly	Gln	Arg	Leu	Leu
				405					410					415	
Phe	Arg	Phe	Met	Lys	Thr	Pro	Asp	Glu	Ile	Met	Ser	Gly	Arg	Thr	Asp
			420					425					430		
Arg	Leu	Glu	His	Leu	Glu	Ser	Gln	Glu	Leu	Asp	Glu	Gln	Ile	Tyr	Gln
		435					440					445			
Glu	Asp	Glu	Cys												
		450													